AD-A120 059 MASSACHUSETTS INST OF TECH CAMBRIDGE F/G 12/1
WHEN IT SEEMS DESIRABLE TO IGNORE DATA.(U)
SEP 82 H CHERNOFF N00014-75-C-0555
UNCLASSIFIED TR-ONR-24 NL

1 of 1
AD A
120059

END
DATE FILMED
11 82
DTIC

WHEN IT SEEMS DESIRABLE TO IGNORE DATA

by

HERMAN CHERNOFF

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

TECHNICAL REPORT NO. ONR 24

SEPTEMBER 1982

PREPARED UNDER CONTRACT

N00014-75-C-0555 (NR-609-001)

FOR THE OFFICE OF NAVAL RESEARCH

STATISTICS CENTER

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

CAMBRIDGE, MASSACHUSETTS

WHEN IT SEEMS DESIRABLE TO IGNORE DATA

by

HERMAN CHERNOFF
Massachusetts Institute of Technology

ABSTRACT

An experiment designed to detect the relative motion of two astronomical objects raised the problem of testing, against shift alternatives, the hypothesis H_o that two energy distributions are equivalent. The relevant data consist of independent Poisson counts X_{ij} with means $\lambda_j p_{ij} T_{ij}$ where λ_j is the intensity of radiation from the j-th object, p_{ij} is the probability that a random photon from the j-th object has energy in a small interval centered about e_i and T_{ij} is the time duration allocated to the count X_{ij}. The hypothesis H_o implies that $p_{i1} = p_{i2}$ for $i = 1, 2, \ldots, m$.

A natural test uses the statistic $\Sigma e_i(\hat{p}_{i2} - \hat{p}_{i1})$ where the $\hat{p}_{ij}$ are estimates of p_{ij}. For intervals where the p_{ij} were anticipated to be small, the experimenter chose small T_{ij} values and hence those $\hat{p}_{ij}$ were highly variable. Consequently, common sense suggests that the corresponding e_i, and X_{ij} be omitted in the above statistic, a practice which may be regarded as sinful by statistical dogma. This issue and others raised by the effects of small T_{ij} lead to the consideration of alternative test statistics and their relative efficiencies as well as the design problem of selecting T_{ij}.

Key Words: Hypothesis testing, optimal design, Pitman efficacy, Poisson
AMS 1980 Subject Classifications: Primary 62F05; Secondary 62G10;

1. Introduction

A satellite based experiment, designed to detect a Doppler effect measuring the relative motion of two neighboring astronomical objects, uses an instrument which can count the number of photons received from either one of these objects within a narrow energy band for a specified duration of time [2]. The total allocated time is divided into [illegible] length intervals, each assigned to a distinct energy band and to one of the astronomical objects. A straightforward analysis of the resulting data raises some puzzling questions which will be addressed here.

The above analysis compares the estimated mean difference of energy levels of the two objects with its estimated standard deviation to see if there is a significant difference. However, since the experimental design allocated relatively little time to energy bands in which the frequency of photons were anticipated to be small, there are some bands with very small or zero counts. With the analysis used it seems preferable, intuitively, to ignore the data in some of these short time, low count intervals. Moreover, with the inclusion of one of these, the change of a single observed zero count to a count of one would have a profound effect on the significance of the results.

Statistical dogma regards the ignoring of data as sinful, yet common sense seems to urge us to commit this sin. The fact that the outcome of an extremely expensive experiment involving hundreds of counts where "the action is", should be greatly affected by the absence or presence of a single count is puzzling. Given another opportunity to repeat this experiment, how should the time intervals be allocated? Most of these issues are addressed in this paper.

This situation does not qualify as a paradox, since procedures that do not make optimal use of the data may perform better by ignoring noisy data liable to be assigned too great a weight. A minor adjustment of the procedure can reduce the effect of the single count, an effect which is, to some extent, a small sample phenomenon. However asymptotic analysis reveals that the main issues are not of a small sample nature and that the "natural" analysis of the data makes considerably less than full use of the available information.

The experiment is described in Section 2. The "natural" analysis and its difficulties together with some mitigating modifications are presented in Section 3. The discussion in Section 4 of a parametric one-sample version of the problem contributes some insights and bounds on efficacy. Finally, in Section 5 several alternative approaches are described and evaluated for the original two-sample problem.

In Section 6, the efficiencies of several approaches are calculated. The significance levels achieved using these methods are presented for four observed data sets. Finally, there is an Appendix where detailed derivations are presented for some of the less obvious results presented in the first five sections.

I wish to thank Joseph Gastwirth for the benefit of some illuminating discussion and the suggestion to use the Mann-Whitney approach discussed in Section 5.

2. Experiment and Notation

Let X_i represent the number of photons observed during the i-th time period of length T_i from a source of intensity λ in a narrow band of energy, centered about e_i, which contributes a proportion p_i to the intensity. Thus we have independent observations $X_1, X_2, \ldots$ where

$$(2.1) \qquad \mathcal{L}(X_i) = \mathcal{P}(\lambda p_i T_i) \qquad i = 1, 2, \ldots, m$$

and where $\mathcal{L}(X)$ represents the distribution of X, $\mathcal{P}(t)$ represents the Poisson distribution with mean t, and $\sum_{i=1}^m p_i = 1$. Let

$$(2.2) \qquad \mu = \sum_{i=1}^{m} p_i e_i$$

be the mean energy of photons from the source (neglecting the effect due to variation of energy within a band).

The parameters λ, p_i, and μ may be estimated by

$$(2.3) \qquad \hat{\lambda} = \sum_{i=1}^{m} X_i T_i^{-1}$$

$$(2.4) \qquad \hat{p}_i = \hat{\lambda}^{-1} X_i T_i^{-1} \quad ,$$

and

$$(2.5) \qquad \hat{\mu} = \sum_{i=1}^{m} \hat{p}_i e_i \quad .$$

Asymptotic analysis indicates that for large λT_i, the distribution of $\hat{\mu}$ is approximately normal with mean μ and variance

$$(2.6) \qquad \sigma^2 = \lambda^{-1}\sum_{i=1}^{m} p_i T_i^{-1}(e_i - \mu)^2 = \lambda^{-1}\Big[\sum_{i=1}^{m} p_i T_i^{-1}\Big]\Big[\sum_{i=1}^{m} w_i(e_i - \mu)^2\Big]$$

where

$$(2.7) \qquad w_i = p_i T_i^{-1}/\Big[\sum_{j=1}^{m} p_j T_j^{-1}\Big] \quad , \qquad i = 1, 2, \ldots, m \quad .$$

For the two sample problem, we introduce X_{i1}, X_{i2}, T_{i1}, T_{i2}, λ_1, λ_2, μ_1, μ_2, p_{i1}, p_{i2}, σ_1^2, σ_2^2, w_{i1} and w_{i2}. To test the hypothesis H_o that there is no difference in energy distributions we apply

$$(2.8) \qquad Z = \frac{\hat{\mu}_2 - \hat{\mu}_1}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}}$$

where σ_j^2 is derived from σ^2 by replacing λ, p_i and μ by $\hat{\lambda}_j$, $\hat{p}_{ij}$ and $\hat{\mu}_j$. Under the null hypothesis of no difference, Z should be approximately normally distributed with mean 0 and varaince 1.

3. Data and Analysis

The experiment consisted of four separate parts, each of which contributed to rejecting the null hypothesis H_o. We present in Table 1, the data from one of the parts which illustrates the problems raised in the introduction. The analysis presented there is based on the use of only the data between $e_i = 2580$ to $e_i = 2660$.

The above analysis is the first performed by the author on this data set. Here those values of i which seemed intuitively unsafe were ignored. It was compared with an independent previous analysis by the experimenters which turned out to be equivalent in formula, but different in that they had included the case $e_o = 2560$. They obtain

$\hat{\lambda}_1 = 0.00538$	$\hat{\mu}_1 = 2601.2$	$\hat{\sigma}_1 = 4.65$	$Z = 3.51$
$\hat{\lambda}_2 = 0.00432$	$\hat{\mu}_2 = 2624.4$	$\hat{\sigma}_2 = 4.68$	

The substantial difference between the two results is due in part to the discrepant weights given to $i = 0$. Indeed, those weights become $\hat{w}_{01} = 0.16$ and $\hat{w}_{02} = 0.00$. Moreoever, if the count of $X_{02} = 0$ is replaced by $X_{02} = 1$, there is another dramatic change with the resulting $Z =$ [illegible]. To a substantial extent these effects can be regarded as small sample effects and they can be reduced considerably by the simple device described below.

Under the null hypothesis H_o, p_{i1} and p_{i2} have a common value p_i. The high variability of the contributions to σ_1^2 and σ_2^2 of the estimates $\hat{p}_{i1}$ and $\hat{p}_{i2}$ would be reduced considerably if the estimates of variance $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ used the simple pooled estimate of p_i.

TABLE 1. Data and Estimates of Parameters and Weights Based on Data for e Ranging from 2580 to 2660 Electron Volts, and a Unit of T_{ij} is 0.32 Seconds

i	e_i	T_{i1}	X_{i1}	T_{i2}	X_{i2}	p_{i1}	p_{i2}	w_{i1}	w_{i2}
	2540	3803.4	3	3.4	0				
0	2560	9017.0	8	1012.4	0				
1	2580	9498.8	12	5831.1	4	0.28	0.16	0.26	0.23
2	2600	9281.8	13	9606.4	4	0.31	0.10	0.29	0.09
3	2620	9132.5	6	9518.8	13	0.15	0.32	0.14	0.28
4	2640	8763.0	7	9281.2	9	0.18	0.22	0.18	0.21
5	2660	5425.4	2	9082.7	8	0.08	0.20	0.13	0.19
	2680	0.0	0	8402.2	3				
	2700	0.0	0	5177.2	3				
	2720	0.0	0	930.8	0				

$\hat{\lambda}_1 = 0.00449$ $\hat{\mu}_1 = 2609.3$ $\hat{\sigma}_1 = 4.43$

$Z = 2.33$

$\hat{\lambda}_2 = 0.00432$ $\hat{\mu}_2 = 2624.4$ $\hat{\sigma}_2 = 4.68$

$$\hat{p}_{i3} = \frac{\hat{\lambda}_1 \hat{p}_{i1} + \hat{\lambda}_2 \hat{p}_{i2}}{\hat{\lambda}_1 + \hat{\lambda}_2} \tag{3.1}$$

Another possibility would be to use the <u>normalized pooled</u> estimates

$$\hat{p}_{i4} = \frac{\hat{q}_i}{\sum_{j=1}^{5} \hat{q}_j} \tag{3.2}$$

where

$$\hat{q}_i = \frac{X_{i1} + X_{i2}}{\hat{\lambda}_1 T_{i1} + \hat{\lambda}_2 T_{i2}} \tag{3.3}$$

In Table 2 we summarize the results where $e_i = 2560$, i.e. $i = 0$, is included and excluded for each of the approaches and for both [illegible] and $X_{02} = 1$ when $i = 0$ is included.

The pooling prescriptions ameliorate considerably one of the problems we observed, but they fail to address the philosophical question of what right we have to omit data. It is clear from Equation (2.6) that it is unwise to include e_i values for which the T_i are <u>relatively</u> small even if these T_i are absolutely large. The problem is not merely a small

TABLE 2.

	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\mu}_2$	$\hat{\sigma}_2$	z
1. Unpooled					
2560 excluded	2609.3	4.43	2624.4	4.68	2.33
2560 included	2601.2	4.65	2624.4	4.68	3.51
2560 included, $X_{02}=1$	2601.2	4.65	2612.4	10.52	0.97
2. Simple Pooled					
2560 excluded	2609.3	5.03	2624.4	5.12	2.09
2560 included	[illegible]	5.25	2624.4	10.51	1.97
2560 included, $X_{02}=1$	2601.2	5.32	2612.4	10.23	0.97
3. Normalized Pooled					
2560 excluded	2609.3	5.14	2624.4	5.19	2.06
2560 included	2601.2	5.41	2624.4	12.45	1.71
2560 included, $X_{02}=1$	2601.2	5.39	2612.4	9.98	0.99

sample problem for which our prescription of pooling or some other ad hoc expedient would be adequate. On the other hand, as indicated in the introduction, it hardly qualifies to be called a paradox, since there never was any claim of optimality, asymptotic or otherwise, for the procedure used. When a suboptimal procedure is used it is not too surprising to find that supressing data, which may be given [illegible] by the procedure, may be desirable.

4. A One-Sample Version of the Problem

While we may not be dealing with a paradox, the theoretical problem still remains of how one should properly analyze the data so that the sin of ignoring data can be avoided. Of course, this does not vitiate the previous analysis. It provides an opportunity to compare that rather straightforward analysis with some sort of optimal alternative to see if there has been a serious loss of efficiency. A further consequence of the solution of the theoretical problem is that it casts light on the problem of the optimal choice of the time intervals T_{ij}.

Let us assume that the p_i are of a prescribed functional form, uniquely determined except for a one dimensional parameter θ representing the velocity of the astronomical object studied. Then we have a problem involving 4 parameters λ_1, θ_1, λ_2 and θ_2 where θ_j determines the proportions $p_{ij} = p_i(\theta_j)$, $j = 1, 2$. For asymptotic considerations we are concerned mainly with the estimate of $\theta_2 - \theta_1$.

We present a more or less standard analysis deriving score functions and information matrices using the likelihood function. For simplicity let us first consider a one-sample version of our problem with parameters λ, θ and data T_i, X_i where $T = \sum T_i$ is assumed to be large. The likelihood function is

$$L = \prod_{i=1}^{m} e^{-\lambda T_i p_i(\theta)} [\lambda T_i p_i(\theta)]^{X_i} / X_i! \quad , \tag{4.1}$$

and

$$\log L = -\lambda \sum_{i=1}^{m} T_i p_i(\theta) + \sum_{i=1}^{m} X_i \log[\lambda T_i p_i(\theta)] - \log X_i!$$

The score function is $\underline{Y} = (Y_1, Y_2)'$ where

$$\begin{aligned} Y_1(\lambda,\theta) &= \frac{\partial \log L}{\partial \lambda} = \lambda^{-1} \sum [X_i - \lambda T_i p_i(\theta)] \\ Y_2(\lambda,\theta) &= \frac{\partial \log L}{\partial \theta} = \sum [X_i - \lambda T_i p_i(\theta)] \frac{\partial \log p_i(\theta)}{\partial \theta} \end{aligned} \tag{4.2}$$

and the Fisher information matrix is

$$J(\lambda,\theta) = E(\underline{Y}\underline{Y}') = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \tag{4.3}$$

where

$$\begin{aligned} \alpha &= \lambda^{-1} \sum T_i p_i(\theta) = \lambda^{-1} \mathcal{E}_\theta(T_i) \\ \beta &= \sum T_i p_i \frac{\partial \log p_i(\theta)}{\partial \theta} = \mathcal{E}_\theta\left[T_i \frac{\partial \log p_i}{\partial \theta}\right] \\ \gamma &= \lambda \sum T_i p_i \left(\frac{\partial \log p_i(\theta)}{\partial \theta}\right)^2 = \lambda \mathcal{E}\left[T_i \left(\frac{\partial \log p_i}{\partial \theta}\right)^2\right] \end{aligned} \tag{4.4}$$

Here $\mathcal{E}_\theta$ is used to represent a formal expectation with respect to the distribution determined by $p_i(\theta)$ and with the index i as the underlying random variable.

If our problem were that of testing $H_o: \theta = \theta_o$, an asymptotically locally optimal procedure would use the test statistic $W_1 = \hat{\theta} - \theta_o$ where $\hat{\theta}$ is the maximum likelihood estimate of θ. This has Pitmann efficacy $[J^{22}(\lambda,\theta_o)]^{-1}$ where $[J^{22}]^{-1} = \gamma - \beta^2/\alpha$ and efficacy is measured per unit time allocated to the experiment. Pitman efficacy is defined in [1]. An asymptotically equivalent test statistic is the second component of $J^{-1}(\hat{\lambda},\theta_o)\underline{Y}(\hat{\lambda},\theta_o)$ where $\hat{\lambda}$ is the maximum likelihood estimate of λ. Note that this statistic is

$$(4.5) \qquad W_2 = [-\hat{\beta}Y_1(\hat{\lambda},\theta_o) + \hat{\alpha}Y_2(\hat{\lambda},\theta_o)]/(\hat{\alpha}\hat{\gamma} - \hat{\beta}^2)$$

where $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are derived from α, β, γ and δ by substituting $(\hat{\lambda},\theta_o)$ for (λ,θ). Moreoever if $\hat{\lambda}$ were replaced by $\hat{\lambda}^*$, the maximum likelihood estimate of λ when θ is assumed to be θ_o, the resulting statistice W_2^* is also asymptotically locally equivalent to W_2 and $\hat{\theta} - \theta_o$. Incidentally, it is easy to see that $\hat{\lambda}^* = \sum X_i/\sum T_i p_i(\theta_o)$ is the solution of $Y_1(\hat{\lambda}^*,\theta_o) = 0$.

If λ were known, one could use $W_3 = Y_2(\lambda,\theta_o)$ with the improved efficacy $T^{-1}J_{22}(\lambda,\theta_o) = T^{-1}\gamma(\lambda,\theta_o)$.

Since W_2 is almost a linear function of the $(X_i - \hat{\lambda}T_i p_i)$ and W_3 is a linear function of the $(X_i - \lambda T_i p_i)$, they are much less sensitive to the variance enhancing effects of small T_i then are expressions which are linear in $X_i T_i^{-1}$. This suggests how information involving X_i for relatively low values of T_i may be incorporated without degrading performance. Thus the issue of ignoring data has been confronted in this *one-sample* problem.

The conclusions in the one-sample problem lead to suggestions for handling the two-sample problem. In brief it seems that shifts in the distribution may be more effectively detected by estimating other linear functions of the p_i than $E(e_i) = \sum p_i e_i = \mu$.

5. The Two-Sample Problem

We return to the two-sample problem. If the $p_i(\theta)$ were known functions of θ, then our problem would be simple. It would suffice to use $W_4 = \hat{\theta}_2 - \hat{\theta}_1$ as a test statistic where θ_j determines the $p_{ij} = p_i(\theta_j)$. Alternatively one could use W_5, the difference of the W_2 statistic for each sample. *Each of these test statistics has optimal efficacy (per unit time) of*

$$\text{(5.1)} \qquad E_0 = ((T_1 + T_2)[\left(\gamma_1 - \frac{\beta_1^2}{\alpha_1}\right)^{-1} + \left(\gamma_2 - \frac{\beta_2^2}{\alpha_2}\right)^{-1}]^{-1}$$

where the subscripts refer to the sample. Incidentally one may attack the design problem of allocating the T_{1i} and T_{2i} to optimize the above efficacy.

The above resolution is inadequate for the original version of the problem where it was not assumed that the $p_i(\theta)$ were well known functions *of the parameter θ. Nevertheless it does suggest the possibility of using a test statistic of the form*

$$\text{(5.2)} \qquad W_6 = \sum [(X_{i2} - X_{i1}) - (\hat{\lambda}_2 T_{i2} - \hat{\lambda}_1 T_{i1}) \left(\frac{X_{i1} + X_{i2}}{\hat{\lambda}_1 T_{i1} + \hat{\lambda}_2 T_{i2}} \right)] a_i$$

An alternative suggestion is to use a statistic of the form

$$\text{(5.3)} \qquad W_6^* = \sum a_i \left[\left(\frac{X_{i2}}{\hat{\lambda}_2} - \frac{X_{i1}}{\hat{\lambda}_1} \right. - (T_{i2} - T_{i1}) \left(\frac{X_{i1} + X_{i2}}{\hat{\lambda}_1 T_{i1} + \hat{\lambda}_2 T_{i2}} \right) \right]$$

since the term which a_i multiplies in (5.3) seems more [illegible] a estimate of $p_{i2} - p_{i1}$. However these two statistics are effectively equivalent. *More precisely*

$$\text{(5.4)} \qquad W_6 = \frac{2\hat{\lambda}_1\hat{\lambda}_2}{\hat{\lambda}_1 + \hat{\lambda}_2} W_6^* + 2\sum a_i \frac{U_i}{V_i}$$

where

$$\hat{V}_i = \hat{\lambda}_1 T_{i1} + \hat{\lambda}_2 T_{i2}$$

$$\text{(5.5)} \qquad V_i = \lambda_1 T_{i1} + \lambda_2 T_{i2}$$

$$U_i = \hat{\lambda}_1 T_{i1} X_{i2} - \hat{\lambda}_2 T_{i2} X_{i1}$$

A second alternative which does not appear to be effectively equivalent to W_6 is to use statistics of the form

$$\text{(5.6)} \qquad W_7 = \sum a_i U_i$$

which would be effectively equivalent to one of the form

$$\text{(5.7)} \qquad W_7^* = \sum a_i(\hat{p}_{i2} - \hat{p}_{i1}) = \sum a_i \left[\frac{X_{i2}}{\hat{\lambda}_2 T_{i2}} - \frac{X_{i1}}{\hat{\lambda}_1 T_{i1}} \right]$$

Note that W_7^* is a minor generalization of $\hat{\mu}_2 - \hat{\mu}_1$.

Still another approach consists of imitating a nonparametric test statistic such as that of Mann-Whitney. Thus we could use an estimate of $P[E_2 > E_1] - P[E_1 > E_2]$ where E_1 and E_2 are the energies of independent random photons from the two astronomical objects. For this we use

(5.8) $$W_8 = \sum_{i<j} \hat{p}_{i1}\hat{p}_{j2} - \sum_{i>j} \hat{p}_{i1}\hat{p}_{j2} \; .$$

It isn't exactly clear what the natural generalization of nonparametric test statistics is in this context. One proposal is to use

(5.9) $$W_9 = \sum_{i,j} \hat{p}_{i1} a_{ij} \hat{p}_{j2}$$

where $A = \{a_{ij}\}$ is a skew symmetric matrix. The statistic W_8 is a special case of W_9 where $a_{ij} = +1$ if $i < j$ and $a_{ij} = -1$ if $i > j$. For testing H_0, an effectively equivalent statistic to W_9 is

(5.10) $$W_9^* = \hat{\lambda}_1\hat{\lambda}_2 W_9 = \sum_{i,j} a_{ij} \frac{X_{i1}X_{i2}}{T_{i1}T_{i2}}$$

To evaluate the efficacies of these statistics we calculate the means and variances of their asymptotic (normal) distributions. These are, for W_6

$$\mu_6 = 2\sum a_i \frac{(p_{i2} - p_{i1})}{(\lambda_1 T_{i1})^{-1} + (\lambda_2 T_{i2})^{-1}}$$

(5.11)

$$\sigma_6^2 = 4\underline{a}'B_6\underline{a}$$

for W_7^*,

(5.12)

$$\mu_7^* = \sum a_i (p_{i2} - p_{i1})$$

$$\sigma_7^{*2} = \underline{a}'B_7^*\underline{a}$$

and for W_9^*

(5.13)

$$\mu_9^* = \lambda_1\lambda_2 \sum_{i,j} p_{i1} a_{ij} p_{j2} = \lambda_1\lambda_2 \sum_{i,j} p_{i1} a_{ij} (p_{j2} - p_{i2})$$

$$\sigma_9^{*2} = \lambda_1^2\lambda_2^2 \sum_j \left[\frac{p_{j1}}{\lambda_1 T_{j1}} \left(\sum_i a_{ij} p_{i2}\right)^2 + \frac{p_{j2}}{\lambda_2 T_{j2}} \left(\sum_i a_{ij} p_{i1}\right)^2 \right]$$

where B_6 and B_7^* are described below.

Let us assume that $p_{ij} = p_i(\theta_j)$ where θ_j is a translation parameter close to θ_0. Then μ_6 and μ_7^* can be approximated by expressions of the form $(\theta_2 - \theta_1)\underline{a}'\underline{\delta}$ and σ_6^2 and σ_7^{*2} by $\underline{a}'B\underline{a}$. For the statistics W_6 and W_7^* the efficacy for testing H_0: $\theta_1 = \theta_2$ is given by $T^{-1}(\underline{a}'\underline{\delta})^2/(\underline{a}'B\underline{a})$. The optimal choice of $\underline{a}$ would be $B^{-1}\underline{\delta}$ and the corresponding efficacy $T^{-1}\underline{\delta}'B^{-1}\underline{\delta}$ if B were nonsingular. However B is singular since there is a vector $\underline{v}$ for which both $B\underline{v} = 0$ and $\underline{\delta}'\underline{v} = 0$. Then $(B+\underline{v}\underline{v}')^{-1}\underline{\delta}$ and $T^{-1}\underline{\delta}'(B+\underline{v}\underline{v}')^{-1}\underline{\delta}$ provide

the optimal $\underline{a}$ and efficacy. The statistic W_9^* must be treated differently. There μ_9^* is approximated by an expression of the form $(\theta_2-\theta_1)\lambda_1\lambda_2\sum q_i p_i(\theta_o)\delta_i$ and σ_9^{*2} by $\lambda_1^2\lambda_2^2\sum q_j^2 p_j(\theta_o)/u_j$, where

$$q_i = \sum_j a_{ij} p_j(\theta_o) \quad . \tag{5.14}$$

The efficacy of W_9^* is $T^{-1}[\sum q_i p_i(\theta_o)\delta_i]^2/[\sum q_j^2 p_j(\theta_o)/u_j]$ and can be maximized easily by the method of Lagrange multipliers when we note that the condition that $\|a_{ij}\|$ be skew symmetric is eqiuivalent to the condition on $\underline{q}$ that $\underline{q}'\underline{p}(\theta_o) = 0$. Then an optimal choice of a_{ij} is such that

$$q_i = \nu_1\delta_i u_i + \nu_2 u_i \tag{5.15}$$

where ν_1 and ν_2 are Lagrange multipliers. More details on these results follow.

For W_6, we have $\underline{\delta}_6$ and B_6, given by

$$\delta_{i6} = 2p_i(\theta_o)\frac{\partial \log p_i(\theta_o)}{\partial\theta_o} u_i \tag{5.16}$$

where

$$u_i = [(\lambda_1 T_{i1})^{-1} + (\lambda_2 T_{i2})^{-1}]^{-1} \quad . \tag{5.17}$$

and

$$B_6 = 4B_6^{**} \tag{5.18}$$

where

$$B_6^{**} = D + s_o\underline{d}\underline{d}' - \underline{d}\underline{p}'(\theta_o) - \underline{p}(\theta_o)\underline{d}' \tag{5.19}$$

with $d_i = p_i(\theta_o)u_i$, $D = \text{diag}(d_i)$ and $s_o = \sum p_i(\theta_o)u_i^{-1}$. Note that $B_6^{**}\underline{v}_6 = 0$ and $\underline{\delta}_6'\underline{v}_6 = 0$ where $v_{i6} = u_i^{-1}$.

For W_7^* we have $\underline{\delta}_7^*$ and B_7^* given by

$$\delta_{i7}^* = p_i(\theta_o)\,\frac{\partial \log p_i(\theta_o)}{\partial\theta} \tag{5.20}$$

and

$$B_7^* = \bar{D} + s_o\underline{p}(\theta_o)\underline{p}(\theta_o)' - \underline{p}(\theta_o)\underline{d}' - \underline{d}\underline{p}'(\theta_o) \tag{5.21}$$

with $\bar{D} = \text{diag}(\bar{d}_i)$, $\bar{d}_i = p_i(\theta_o)u_i^{-1}$ and $s_o = \sum \bar{d}_i$ is as above. Here also we note that $B_7^*\underline{v}_7 = 0$ and $\underline{\delta}_7^*\underline{v}_7 = 0$ for $v_{i7} = 1$.

For W_9^* we have $\delta_{i9} = -\partial\log p_i(\theta_o)/\partial\theta$ and $\underline{u}$ is as above. Applying the method of Lagrange multipliers, we derive

$$v_1^{-1} = \mathcal{E}_{\theta_o}\{\beta_{i9}^2 u_i\} - \frac{[\mathcal{E}_{\theta_o}\{\beta_{i9} u_i\}]^2}{\mathcal{E}_{\theta_o}(u_i)} \tag{5.22}$$

$$v_2 = - \, {}_1\mathcal{E}_{\theta_o}\{\beta_{i9} u_i\}/\mathcal{E}_{\theta_o}\{u_i\} \tag{5.23}$$

and the maximum value of the efficacy is $T^{-1} v_1^{-1}$.

For each of the statistics W_6, W_7^* and W_9^*, the best choice of coefficients a_i depends on knowledge of $p_i(\theta)$. While an incorrectly specified model for $p_i(\theta)$ will lead to a suboptimal choice of coefficients, the resulting statistic can still be used to test the hypothesis H_o under which the mean is zero and the variance may be estimated by using the pooled estimates of $p_{i1} = p_{i2}$.

In the meantime one may compute the relative efficiencies of these statistics under the *assumption that* $p_i(\theta)$ is of a special form. In particular, for our special problem, the forms of interest are those where the underlying density is normal, or a mixture of two or three normals, and p_{ij} is the integral of this density over the range $e_i + \theta_j \pm h/2$ with $h = e_{i+1} - e_i$.

Several problems have been neglected and remain somewhat unresolved. These are due to two facts. First the energy bands of width h are not very narrow, and second, there is a truncation effect since there may be

substantial radiation outside of the bands to which time has been allocated. Thus, while deviations from H_o may be detected by observing, for large T, that $p_{i1} \neq p_{i2}$, the extent of the translation can not easily be determined without assuming a functional form for the density of the energy. Nor is it possible to check that the failure of H_o corresponds to a translation of the energy distribution.

This problem is relatively minor if there are many narrow energy bands and if there is little density in those bands near the boundary of the region to which substantial time T_{ij} has been allocated. Otherwise, even $\sum(p_{i2} - p_{i1})e_i$ is likely to be a poor estimate of the shift. If the bands are wide e_i may not accurately represent the average energy in the band. If there is substantial density at the boundaries of the region studied, the probabilities being estimated are *really* conditional probabilities given that the energy is in the prescribed region. Then a translation shift in the energy distribution does not correspond exactly to a translation shift in the conditional distribution.

If these difficulties were negligible, one could estimate the translation parameter by seeing how much of shift h from $p_{i+h,2}$ to p_{i2}^* would be required for one of the resulting statistics W_6, W_7^*, W_9^* to cross zero.

Another question that has not been carefully examined is the relevance of Pitman efficiency, a measure designed for dealing with small shifts, if the shift is substantial compared to the standard deviations of the components of the density corresponding to the modes or spectral lines studied.

6. Tables of Efficacies and Significance

The experiment yielded four data sets, the fourth of which appears in Table 1. The others are presented in Table 3. These data suggest models for $p_i(\theta)$ of the form

$$p_i(\theta) = \int_{i-1/2}^{i+1/2} f(x - \theta)dx$$

where

$$f(x) = \sum_{j=j_1}^{j_2} \sigma_j^{-1}\phi(x/\sigma_j)dx$$

and ϕ is the standard normal density. These models were used for our theoretical evaluations. This method of choosing models, loosely fitting the data, may bias our results to yield apparent efficiences somewhat larger than deserved for methods whose coefficients are "tuned" to optimize with respect to these "fitted" models. The parameters of the models are presented in Table 4. The efficacies of various methods are presented in Table 5. These methods are applied to the data sets, and in Table 6, the corresponding Z values are presented. Since the significance level or P values are given approximately by

$$P = \Phi(Z)$$

where Φ is the standard normal c.d.f., these levels are not presented explicitly. More details follow.

Table 3. Three Data Sets

$e_i = e_0 + b_i$

1. $e_o = 890$ $b = 4$

i	T_{i1}	X_{i1}	T_{i2}	X_{i2}
1	1319.4	2	317.6	0
2	6032.0	3	3119.4	1
3	7104.5	3	4134.6	2
4	7346.2	9	4091.7	2
5	7304.9	6	4844.9	1
6	7226.7	14	6692.7	2
7	7041.7	12	7268.8	12
8	5990.5	14	7120.5	13
9	4252.2	4	7266.9	7
10	3160.1	9	7201.5	15
11	3051.4	9	7743.9	15
12	2259.6	2	6084.3	13
13	1180.6	0	4131.4	4
14	226.3	0	3099.8	2

2. $e_o = 2365$ $b = 10$

i	T_{i1}	X_{i1}	T_{i2}	X_{i2}
1	2356.6	1	279.8	0
2	2278.3	2	1853.1	0
3	2538.2	6	2183.3	2
4	3940.7	8	2214.6	0
5	4857.1	9	2275.5	4
6	4717.4	14	2340.3	4
7	4553.1	18	3407.7	7
8	3683.3	14	4533.7	14
9	2840.8	5	4652.6	16
10	2522.2	3	4714.0	9
11	2473.3	5	4435.7	5
12	2478.5	3	3336.3	4
13	2165.9	1	2688.6	1
14	1132.1	2	2515.2	0
15	283.6	0	2487.2	0

3. $e_o = 890$ $b = 4$

i	T_{i1}	X_{i1}	T_{i2}	X_{i2}
1	1278.7	0	0.0	0
2	5775.5	6	6354.0	0
3	6736.0	6	10228.6	7
4	9985.8	6	13426.2	6
5	12132.8	15	13204.6	8
6	12023.2	19	13045.5	14
7	11976.4	30	13053.6	14
8	12025.4	28	12812.3	23
9	11867.2	18	12679.2	11
10	11710.0	20	12595.6	25
11	10082.1	14	10808.8	20
12	7562.2	3	7519.8	20
13	6296.0	5	6703.6	6
14	1581.0	1		6

Table 4. Parameters of Three Models for $p_i(\theta)$

Model 1				Model 2				Model 3			
n	λ_1	λ_2		n	λ_1	λ_2		n	λ_1	λ_2	
3	0.0165	0.0135		1	0.0263	0.0275		1	0.0045	0.0043	
j	μ_j	σ_j	w_j	j	μ_j	σ_j	w_j	j	μ_j	σ_j	w_j
1	3.5	1.6	0.15	1	8.0	2.5	1.0	1	4.0	2.0	1.0
2	7.5	1.6	0.36								
3	11.0	1.6	0.49								

The efficacies presented in Table 5 are for the statistics testing $H_o: p_{i1} = p_{i2}$, $1 \leq i \leq m$ against local shift alternatives to H_o^*: the model applies with $\theta_1 = \theta_2$ (approximately 0).

1. E_o is the efficacy of the optimal parametric test.
2. E_{06}, E_{07}, and E_{09} are the efficacies of W_6, W_7, and W_9 using the optimal coefficients
3. E_{17}^* is the efficacy of the "natural test" based on the statistic $W = \Sigma e_i(\dot{p}_{i2} - \dot{p}_{i1})$ or its equivalent $W_{17}^* = \Sigma i(\dot{p}_{i2} - \dot{p}_{i1})$. $(e_i = e_o + bi)$.
4. E_8 is the efficacy of W_8, the Mann-Whitney version of W_9.

For each model, the data set is relevant only in the values of T_{i1} and T_{i2} (design parameters) used, and the observed counts X_{ij} are not relevant. For each model and data set combination, we consider several alternative subsets of the available counts for inclusion in the analysis. Thus if we take $i_1 \leq i \leq i_2$, $T = \sum_{i_1}^{i_2}(T_{i1} + T_{i2})$ and TE is of interest as well as E.

Computations show that $E_{06} = E_{07} = E_{09}$ seems to be true in general. This is not very surprising and should not be too difficult to establish. In particular the equality of E_{06} and E_{07} was anticipated.

Clearly we should have $E_0 > E_{06}$. Possibly because of the loose fitting of the models to the observed data, E_{06} is very close to E_0. On the other hand W_{17}^* is sometimes poor and sometimes very sensitive to the choise of data to be included in the analysis. Generally W_8 does better than W_{17}^* and is less sensitive to the choice of data to be included.

In Table 6 we present the values of $Z = W/\hat{\sigma}_W$ corresponding to various estimators applied to the actual data sets. For each estimator two Z values

Table 5. **Efficacies of Tests for** $H_0:\{p_{i1} = p_{i2}, 1 < i < m\}$

$$\bar{E} = 10^5 * E,\ \bar{T} = 10^{-5} * T$$

d	m	i_1	i_2	$\bar{E}_o$	$\bar{E}_{06}$	$\bar{E}^*_{17}$	$\bar{E}_8$	$\bar{T}$	TE_{06}	TE^*_{17}
1	1	2	13	1.67	1.55	1.24	1.32	1.31	2.03	1.62
1	1	1	14	1.78	1.55	0.87	1.05	1.36	2.11	1.18
1	1	3	12	1.20	1.15	0.78	0.88	[illegible]	1.34	0.91
2	2	2	14	7.01	6.80	6.53	6.77	0.81	5.51	5.29
2	2	1	15	6.68	6.40	5.22	6.04	0.87	5.57	4.54
2	2	1	14	6.85	6.60	5.80	6.39	0.84	5.54	4.87
2	2	3	13	7.31	7.10	6.99	7.08	0.74	5.25	5.17
3	1	3	13	2.10	2.09	1.39	1.59	2.38	4.97	3.31
3	1	3	14	2.33	2.23	1.55	1.75	2.46	5.49	3.81
3	1	4	12	1.49	1.48	1.01	1.13	2.11	3.12	2.13
4	3	2	7	1.92	1.67	1.12	1.11	0.95	1.59	1.06
4	3	1	7	[illegible]	1.49	0.01	0.01	0.99	1.47	0.01
4	3	3	6	1.73	1.71	1.69	1.68	0.71	1.21	1.20
4	3	3	7	1.92	1.87	1.84	1.84	0.85	1.59	1.56

d = data set number
m = model number

Table 6. **Significance Levels and Estimated Shifts**

Z values corresponding to various test statistics and data sets. $P = \Phi(-Z)$ where Φ is the standard normal c.d.f. Data set 5 is data set 4 with X_{22} replaced by 1.

S values are estimated shifts in energy distribution normalized by dividing by energy.

d	m	i_1	i_2	Z_6	Z^*_{07}	Z^*_{17}	Z_8	Z^*_{09}	Z^*_{19}	S^*_{07}	S^*_{17}
1	1	2	13	2.6	2.6	2.2	2.4	3.0	2.9	0.0048	0.0049
1	1	1	14	2.8	2.8	1.8	2.2	2.8	3.3	0.0064	0.0066
1	1	3	12	2.3	2.3	1.8	1.9	3.1	2.1	0.0045	[illegible]
2	2	2	14	1.4	1.5	1.0	1.4	1.1	0.5	0.0049	0.0025
2	2	1	15	1.5	1.5	1.1	1.4	1.5	0.5	0.0040	0.0030
2	2	1	14	1.5	1.5	1.1	1.4	1.1	[illegible]	0.0052	[illegible]
2	2	3	13	1.9	1.9	1.8	2.0	1.8	1.9	0.0059	[illegible]
3	1	3	13	2.9	2.9	2.0	2.3	2.6	3.1	0.0094	[illegible]
3	1	3	14	2.7	2.7	1.8	2.1	2.0	3.1	0.0100	[illegible]
3	1	4	12	3.8	3.8	3.4	3.5	3.6	5.9	[illegible]	[illegible]
4	3	2	7	2.6	2.6	2.3	[illegible]	2.8	1.9	[illegible]	[illegible]
4	3	1	7	1.3	1.3	0.2	0.2	2.7	1.9	0.0156	[illegible]
4	3	3	6	1.9	1.9	1.8	1.8	2.0	1.8	0.0271	[illegible]
4	3	3	7	2.3	2.3	2.2	2.3	2.3	1.7	0.0154	[illegible]
5	3	2	7	2.0	2.0	1.0	1.0	2.1	1.5	0.0204	[illegible]
5	3	1	7	1.3	1.4	0.2	0.2	2.0	1.6	0.0154	[illegible]

d = data set number
m = model number

were calculated corresponding to the use, respectively of $\hat{p}_{3i}$ and $\hat{p}_{4i}$, in the estimation of the standard deviation of the W statistic. Since these Z values were almost always very close to one another, only those corresponding to $\hat{p}_{3i}$ are presented. For W_9, three versions of Z were calculated corresponding to the related statistics W^*_{09}, W^*_{193} and W^*_{194}. The first, W^*_{09} uses the coefficients a_{ij0} selected so that

$$\sum_j a_{ij0} p_j(0) = q_{i0}$$

where q_{i0} are the optimal values of q_i according to the appropriate model. Since the choice of the a_{ij0} is not unique, they were selected so that $a_{i,i+1,0} = a_{i,i+2,0} = \ldots$. The statistic W^*_{193} uses similar coefficents subject to

$$\sum_j a_{ij3} \hat{p}_{j3} = q_{i0}$$

and W^*_{194} is derived similarly. We present Z values for W^*_{09} and W^*_{193}, i.e. Z^*_{09} and Z^*_{19}. Note that Table 6 involves the model and i_1 and i_2 since these determine the coefficients of W_6, W_7, and the W_9 statistics.

Finally, estimates of the mean shift $\theta_2 - \theta_1$ were calculated. To be more specific $\hat{\mu}_2 - \hat{\mu}_1 = bW^*_{17}$ estimates $\mu_2 - \mu_1 = \sum e_i(p_{i2} - p_{i1})$, the shift in the distribution of energy. The relative velocity of the two astronomical objects is approximately proportional to $(\mu_2 - \mu_1)/e_o$ which is estimated by $S^*_{17} = bW^*_{17}/e_o$. To estimate this same parameter using W^*_{07} the following coarse technique was used. Compute

$$W^{*+}_{07} = \sum_{i=i_1}^{i_2-1} \frac{a_{oi}+a_{o,i+1}}{2} \left[\frac{\hat{p}_{i+1,2}}{1-\hat{p}_{i_1,2}} - \frac{\hat{p}_{i,1}}{1-\hat{p}_{i_2,1}} \right]$$

to represent the value of W^*_{07} if θ_1 and θ_2 were each shifted by 1/2 in opposite directions. We approximate s, the number of intervals by which $\theta_2 - \theta_1$ must be shifted, to make the (shifted) W^*_{07} statistic zero by

$$W^*_{07} + s(W^{*+}_{07} - W^*_{07}) = 0$$

and let the corresponding velocity be estimated by the normalized shift

$$S^*_{07} = bs/e_o = bW^*_{07}/e_o(W^*_{07} - W^{*+}_{07})$$

A coarse estimate of the coefficient of variation of $S^*_{07}(S^*_{17})$ is $1/Z^*_{07}(1/Z^*_{17})$.

These estimates are unreliable because of the width of the cells, the truncation effects, and the likelihood that the true shifts are not small. The estimates S^*_{07} and S^*_{17} are quite variable with S^*_{17} generally substantially larger than S^*_{07}.

The careful reader may note that there is a nontrivial difference between the Z values in Tables 2 and 6 corresponding to data set 4. This difference is due to the fact that in Table 2, Equation 2.6 was applied with $\hat{\mu}_1 = \sum e_i \hat{p}_{i1}$ substituting for μ when j = 1 and $\hat{\mu}_2 = \sum e_i \hat{p}_{i2}$ when j = 2. On the other hand for the calculation in Table 6, $\hat{\mu}_3 = \sum e_i \hat{p}_{i3}$ was substituted for μ in both cases.

BIBLIOGRAPHY

[1] Lehmann, E.L., (1975). Nonparametrics: Statistical Methods Based on Ranks, Holden-Day, San Francisco.

[2] Markert, T.H., Canizares, C.R., Clark, G.W., and Winkler, P.F., "High Velocity Asymmetric Doppler Shifts of the X-Ray Emission Lines of Cassiopeia A", to be submitted to Astrophysical Journal.

A. APPENDIX

A.1 Equation (2.6)

$$\hat{\mu} = \sum \hat{p}_i e_i = \frac{\sum X_i T_i^{-1} e_i}{\sum X_i T_i^{-1}} = \frac{N}{D}$$

Using the δ method, the (asymptotic) variance of $\hat{\mu}$, (as $T = \sum T_i \to \infty$) is the variance of

$$Y = \frac{N}{E(D)} - \frac{E(N)D}{[E(D)]^2}$$

Since $E(N) = \lambda\mu$ and $E(D) = \lambda$,

$$Y = \lambda^{-1}[\sum X_i T_i^{-1}(e_i - \mu)]$$

with variance

$$\sigma_Y^2 = \lambda^{-2}\sum \lambda p_i T_i T_i^{-2}(e_i - \mu)^2 = \lambda^{-1}\sum p_i T_i^{-1}(e_i - \mu)^2$$

A.2 Efficacy of $\hat{\theta} - \theta_o$, W_2, W_2^*

(a) $\hat{\theta} - \theta_o$. For θ close to θ_o

(A2.1) $$\mathcal{L}[(\hat{\theta} - \theta)\sqrt{T}] = N(0, TJ^{22}(\lambda, \theta_o))$$

Hence the efficacy of $\hat{\theta} - \theta_o$ (per unit time) is

(A2.2) $$[TJ^{22}(\lambda, \theta_o)]^{-1}$$

[dividing the square of the derivative of the mean by the variance]

(b) W_2.

Let $\underline{\phi} = (\lambda, \theta)'$ with $\theta - \theta_o = O(T^{-1/2})$

(A2.3) $$\underline{W} = J^{-1}(\hat{\lambda}, \theta_o)\underline{Y}(\hat{\lambda}, \theta_o)$$

where $\hat{\lambda} - \lambda = O_p(T^{-1/2})$, $T^{-1}J$ and TJ^{-1} are $O_p(1)$, and $\mathcal{L}[T^{-1/2}\underline{Y}(\lambda, \theta)] = N(0, T^{-1}J)$. Then

(A2.4) $$J^{-1}(\hat{\lambda}, \theta_o) = J^{-1}(\lambda, \theta) + O_p(T^{-3/2}) = J^{-1}(\lambda, \theta_o) + O_p(T^{-3/2})$$

(A2.5) $$\underline{Y}(\hat{\lambda}, \theta_o) = \underline{Y}(\lambda, \theta) + \frac{\partial \underline{Y}(\lambda, \theta)}{\partial \phi} \begin{pmatrix} \hat{\lambda} - \lambda \\ \theta_o - \theta \end{pmatrix} + \cdots$$

$$\frac{\partial \underline{Y}}{\partial \phi} = -J(\lambda, \theta)[1 + o_p(1)]$$

$$\underline{Y}(\hat{\lambda}, \theta_o) = \underline{Y}(\lambda, \theta) - J(\lambda, \theta) \begin{pmatrix} \hat{\lambda} - \lambda \\ \theta_o - \theta \end{pmatrix} + o_p(T^{1/2})$$

(A2.6) $$\underline{W} = J^{-1}(\lambda, \theta)\, \underline{Y}(\lambda, \theta) - \begin{pmatrix} \hat{\lambda} - \lambda \\ \theta_o - \theta \end{pmatrix} + o_p(T^{-1/2})$$

(A2.7) $$\mathcal{L}\{T^{1/2}[W_2 - (\theta - \theta_o)]\} = N(0, TJ^{2\cdot}JJ^{\cdot 2}) = N(0, TJ^{22})$$

where the argument of J can be taken to be (λ, θ_o). It follows that the efficacy of W_2 is $[TJ^{22}(\lambda, \theta_o)]^{-1}$

(c) W_2^*.

We shall show that $\hat{\lambda}^* - \lambda = O_p(T^{-1/2})$ after which the argument used for W_2 can be reproduced without change. The key reason is that $\hat{\lambda} - \lambda$ appears in the first but not in the second component of $\underline{W}$ in (A2.6).

The fact that $\hat{\lambda}^* - \lambda = O_p(T^{-1/2})$ follows directly from the expression for $\hat{\lambda}^*$. A more general approach, not confined to this particular problem, follows

$$0 = Y_1(\lambda_*, \theta_o) = Y_1(\lambda, \theta) + \frac{\partial Y_1(\lambda, \theta)}{\partial \phi} \begin{pmatrix} \lambda_* - \lambda \\ \theta_o - \theta \end{pmatrix} + \cdots$$

(A2.8) $$\lambda_* - \lambda = \frac{Y_1(\lambda, \theta) + \frac{\partial Y_1}{\partial \theta}(\lambda, \theta)(\theta_o - \theta)}{-\frac{\partial Y_1(\lambda, \theta)}{\partial \lambda}}$$

Using the facts that $Y_1(\lambda,\theta) = O_p(T^{\frac{1}{2}})$, $\theta - \theta_o = O(T^{-\frac{1}{2}})$, $\frac{\partial Y_1(\lambda,\theta)}{\partial\theta} = O_p(T)$ and $\frac{\partial Y_1(\lambda,\theta)}{\partial\lambda} = J_{22}(\lambda,\theta)$ which is of the order of magnitude of T, the desired result follows.

A.3 The asymptotic distributions of W_6, W_7^* and W_9^*.

(a) The statistics W_6, W_7^* and W_9^* are clearly asymptotically normal. We shall calculate the means and variances of the approximating distributions by applying the δ method which consists of expanding in terms of $X_{j1} - \lambda_1 P_{j1} T_{j1}$ and $X_{j2} - \lambda_2 P_{j2} T_{j2}$. We use tildes to represent statistics evaluated at $(X_{j1}, X_{j2}) = (\lambda_1 P_{j1} T_{j1}, \lambda_2 P_{j2} T_{j2})$. Note also that

$$W_6^{**} = \frac{1}{2} W_6 = \sum a_j U_j / V_j$$

where

$$\tilde{U}_j = \lambda_1 \lambda_2 T_{j1} T_{j2} (p_{j2} - p_{j1})$$
$$\tilde{V}_j = V_j = \lambda_1 T_{j1} + \lambda_2 T_{j2} \quad .$$

Then the mean for W_6^{**} is

$$\mu_6^{**} = \tilde{W}_6^{**} = \sum a_j \frac{p_{j2} - p_{j1}}{(\lambda_1 T_{j1})^{-1} + (\lambda_2 T_{j2})^{-1}} \quad .$$

and the variance is

$$\sigma_6^{**2} = \sum_j \mathrm{Var}\,(X_{j1}) \left[\frac{\partial \tilde{W}_6^{**}}{\partial X_{j1}} \right]^2 + \mathrm{Var}(X_{j2}) \left[\frac{\partial \tilde{W}_6^{**}}{\partial X_{j2}} \right]^2$$

$$\frac{\partial \widetilde{W}_6^{**}}{\partial X_{i1}} = -a_i \cdot \frac{\lambda_2 T_{i1} T_{i2}}{V_i T_{i1}} + \sum_j a_j \frac{\lambda_2 T_{j1} T_{j2}}{V_j T_{j1}} \cdot \frac{\lambda_1 T_{j1} p_{j1} + \lambda_2 T_{j2} T_{pj2}}{V_j}$$

$$\frac{\partial \widetilde{W}_6^{**}}{\partial X_{i2}} = a_i \cdot \frac{\lambda_1 T_{i1} T_{i2}}{V_i T_{i2}} - \sum_j a_j \frac{\lambda_1 T_{j1} T_{j2}}{V_j T_{j2}} \cdot \frac{\lambda_1 T_{j1} p_{j1} + \lambda_2 T_{j2} p_{j2}}{V_j}$$

$$\sigma_6^{**2} = \sum_i \frac{\lambda_2^2}{T_{i1}^2} \left[-a_i b_i + \sum_j a_j b_j c_j\right]^2 \lambda_1 T_{i1} p_{i1}$$

$$+ \frac{\lambda_1^2}{T_{i2}^2} \left[a_i b_i - \sum_j a_j b_j c_j\right]^2 \lambda_2 T_{i2} p_{i2}$$

where $b_i = T_{i1} T_{i2} / V_i = u_i / \lambda_1 \lambda_2$ and $c_i = (\lambda_1 T_{i1} p_{i1} + \lambda_2 T_{i2} p_{i2}) / V_i$

$$\sigma_6^{**2} = \lambda_1^2 \lambda_2^2 \sum_i \left[\frac{p_{i1}}{\lambda_1 T_{i1}} + \frac{p_{i2}}{\lambda_2 T_{i2}}\right] \left[a_i b_i - \sum_j a_j b_j c_j\right]^2$$

$$= \lambda_1^2 \lambda_2^2 \sum_i s_i \left[a_i b_i - \sum_j a_j b_j c_j\right]^2 = \underline{a}' B_6^{**} \underline{a}$$

where $s_i = p_{i1} (\lambda_1 T_{i1})^{-1} + p_{i2} (\lambda_2 T_{i2})^{-1}$. Let $d_i = s_i u_i^2$, $r_i = c_i u_i$, $t_i = s_i u_i$, and $s_o = \sum s_i$. Then

$$\underline{a}' B_6^{**} \underline{a} = \sum_i d_i a_i^2 + \sum_j s_j \left[-2 a_j b_j (\underline{a}'\underline{r}) + (\underline{a}'\underline{r})^2\right]$$

$$= \sum_i d_i a_i^2 + s_o (\underline{a}'\underline{r})^2 - 2 (\underline{a}'\underline{t}) (\underline{a}'\underline{r})$$

$$B_6^{**} = D_6 + s_o (\underline{r}\underline{r}') - (\underline{r}\underline{t}') - (\underline{t}\underline{r}') = D_6 - C_6 A_6^{-1} C_6'$$

where

$$D_6 = \operatorname{diag}(\underline{d}) ,$$

$$C_6 = (\underline{r}, \underline{t})$$

$$A_6^{-1} = \begin{pmatrix} -s_o & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad A_6 = \begin{pmatrix} 0 & 1 \\ 1 & s_o \end{pmatrix} .$$

If $p_{i1} = p_{i2} = p_i$, $d_i = p_i u_i$, $r_i = d_i$, $t_i = p_i$, and $s_o = \sum p_i u_i^{-1}$.

(b)

$$W_7^* = \sum a_i (\dot{p}_{i2} - \dot{p}_{i1}) = \sum a_i \left(\frac{X_{i2}}{\lambda_2 T_{i2}} - \frac{X_{i1}}{\lambda_1 T_{i1}}\right)$$

$$\mu_7^* = \bar{W}_7^* = \sum a_j (p_{j2} - p_{j1})$$

$$\frac{\partial \widetilde{(\dot{p}_{j2} - \dot{p}_{j1})}}{\partial X_{i1}} = \frac{-\delta_{ij}}{\lambda_1 T_{i1}} + \frac{\lambda_1 T_{j1} p_{j1}}{\lambda_1^2 T_{j1}} \cdot \frac{1}{T_{i1}} = \frac{1}{\lambda_1 T_{i1}} (p_{j1} - \delta_{ij})$$

$$\frac{\partial \widetilde{(\dot{p}_{j2} - \dot{p}_{j1})}}{\partial X_{i2}} = \frac{-1}{\lambda_2 T_{i2}} (p_{j2} - \delta_{ij})$$

$$\sigma_7^{*2} = \sum_i \left\{ \lambda_1 p_{i1} T_{i1} \frac{[a_i - \sum_j p_{j1} a_j]^2}{\lambda_1^2 T_{i1}^2} + \lambda_2 p_{i2} T_{i2} \frac{[a_i - \sum_j p_{j2} a_j]^2}{\lambda_2^2 T_{i2}^2} \right\}$$

$$= \underline{a}' B_7 \underline{a} = \underline{a}' D_7 \underline{a} - 2[\underline{a}' \underline{b}^{(1)} \underline{p}^{(1)\prime} \underline{a} + \underline{a}' \underline{b}^{(2)} \underline{p}^{(2)\prime} \underline{a}]$$

$$+ s_{01} \underline{a}' \underline{p}^{(1)} \underline{p}^{(1)\prime} \underline{a} + s_{02} \underline{a}' \underline{p}^{(2)} \underline{p}^{(2)\prime} \underline{a}$$

where $D_7 = \mathrm{diag}(\underline{s})$, $s_{01} = \sum (\lambda_1 T_{i1})^{-1} p_{i1}$, $s_{02} = \sum (\lambda_2 T_{i2})^{-1} p_{i2}$, $\underline{p}_i^{(j)} = p_{ij}$, and $b_i^{(j)} = (\lambda_j T_{ij})^{-1} p_{ij}$. If $p_{i1} = p_{i2} = p_i$

$$B_7 = D_7 - C_7 A_7^{-1} C_7'$$

where $D_7 = \mathrm{diag}(\bar{\underline{d}})$, $\bar{d}_i = p_i u_i^{-1}$

$$C_7 = (\underline{p}, \underline{d})$$

$$A_7^{-1} = \begin{pmatrix} -s_o & 1 \\ 1 & 0 \end{pmatrix}, \qquad A_7 = A_6 = \begin{pmatrix} 0 & 1 \\ 1 & s_o \end{pmatrix}$$

and

$$\varepsilon_o = \sum p_i u_i^{-1} .$$

(c)

$$W_9 = \dot{\underline{p}}_1' A \dot{\underline{p}}_2$$

where A is skew symmetric.

$$W_9^* = \lambda_1 \lambda_2 W_9 = \sum a_{ij} \frac{X_{i1}}{T_{i1}} \frac{X_{j2}}{T_{j2}}$$

$$u_9^* = \bar{W}_9^* = \lambda_1 \lambda_2 \underline{p}^{(1)\prime} A \underline{p}^{(2)} = \lambda_1 \lambda_2 [(\underline{p}^{(1)} - \underline{p}^{(2)})' A \underline{p}^{(2)}].$$

$$\frac{\widetilde{\partial W_9^*}}{\partial X_{i1}} = \sum_k a_{ik} \frac{\lambda_2 T_{k2} p_{k2}}{T_{i1} T_{k2}} = \frac{\lambda_2}{T_{i1}} \sum_k a_{ik} p_{k2}$$

$$\frac{\widetilde{\partial W_9^*}}{\partial X_{i2}} = -\sum_k a_{ik} \frac{\lambda_1 p_{k1}}{T_{i2}} = -\frac{\lambda_1}{T_{i2}} \sum_k a_{ik} p_{k1}$$

$$\sigma_9^{*2} = \sum_i \left\{ \lambda_1 T_{i1} p_{i1} \left[\frac{\lambda_2}{T_{i1}} \sum a_{ik} p_{k2} \right]^2 + \lambda_2 T_{i2} p_{i2} \left[\frac{\lambda_1}{T_{i2}} \sum a_{ik} p_{k1} \right]^2 \right.$$

If $p_{i1} = p_{i2} = p_i$, let

$$q_i = \sum_j a_{ij} p_j$$

and then

$$\sigma_9^{*2} = \lambda_1^2 \lambda_2^2 \sum_i p_i [(\lambda_1 T_{i1})^{-1} + (\lambda_2 T_{i2})^{-1}] q_i^2 = \lambda_1^2 \lambda_2^2 \sum_i p_i u_i^{-1} q_i^2 .$$

Note that if $\underline{p}^{(1)}$ and $\underline{p}^{(2)}$ are close to $\underline{p}$

$$\mu_9^* = \lambda_1\lambda_2 \sum_i (p_i^{(1)} - p_i^{(2)}) q_i \ .$$

If $\underline{p}^{(j)} = p(\theta_j)$, θ_1, θ_2 close to θ_o, $p_i^{(1)} - p_i^{(2)} = (\theta_1 - \theta_2) p_i(\theta_o) \dfrac{\partial \log p_i(\theta_o)}{\partial\theta}$

$= (\theta_2 - \theta_1)\, \delta_{i9} p_i(\theta_o)$ where $\delta_{i9} = -\partial \log p_i(\theta_o)/\partial\theta$.

A.4 Two maximization problems.

(a) The maximum value of $(\underline{a}'\underline{\delta})^2/(\underline{a}'B\underline{a})$ is obtained by minimizing $\underline{a}'B\underline{a}$ subject to $\underline{a}'\underline{\delta} = K$. Applying the method of Lagrange multipliers

$$B\underline{a} = \nu\underline{\delta}$$
$$\underline{a} = \nu B^{-1}\underline{\delta}$$

and

$$(\underline{a}'\underline{\delta})^2/(\underline{a}'B\underline{a}) = \underline{\delta}'B^{-1}\underline{\delta}$$

provided B is nonsingular. However, in our application $B\underline{v} = 0$ and $\underline{\delta}'\underline{v} = 0$. If we replace B by $F = B + \underline{v}\,\underline{v}'$ and $\underline{a}$ by $\underline{a} + h\,\underline{v}$ with $\underline{a}'\underline{v} = 0$, then

$$(\underline{a} + h\,\underline{v})'\underline{\delta} = \underline{a}'\underline{\delta}$$

and

$$(\underline{a} + h\,\underline{v})'\, F(\underline{a} + h\,\underline{v}) = \underline{a}'B\underline{a} + h^2(\underline{v}'\underline{v})^2 .$$

It is clear that the minimizing $\underline{a} + h\,\underline{v}$ for the new problem coincides with that of the original problem. Thus the maximizing value of $\underline{a}$ is $F^{-1}\underline{\delta}$ and the maximum value of $(\underline{a}'\underline{\delta})^2/(\underline{a}'B\underline{a})$ is $\underline{\delta}'F^{-1}\underline{\delta}$, provided that F is nonsingular.

(b) To minimize $\sum p_i u_i^{-1} q_i^2$ subject to $\sum p_i \delta_i q_i = K$ where $q_i = \sum a_{ij} p_j$ under the condition that $|a_{ij}|$ be skew symmetric. The condition of skew symmetry implies $\sum p_i q_i = 0$. Moreover, given any vector for which $\underline{p}'\underline{q} = 0$, there is a skew symmetric A for which $\underline{q} = A\underline{p}$. Hence, for our minimization problem, skew symmetry is equivalent to $\underline{p}'\underline{q} = 0$, and we may apply the method of Lagrange multipliers to minimize with respect to q.

$$p_i u_i^{-1} q_i = \nu_1 p_i \delta_i + \nu_2 p_i$$

$$q_i = \nu_1 \delta_i u_i + \nu_2 u_i$$

The restrictions yield

$$\nu_1 \sum p_i \delta_i u_i + \nu_2 \sum p_i u_i = 0$$

$$\nu_1 \sum p_i \delta_i^2 u_i + \nu_2 \sum p_i \delta_i u_i = K$$

$$\nu_2 = -\nu_1 \mathcal{E}(\delta_i u_i)/\mathcal{E}(u_i)$$

$$\nu_1 = \frac{K\mathcal{E}(u_i)}{\mathcal{E}(u_i)\mathcal{E}(\delta_i^2 u_i) - [\mathcal{E}(\delta_i u_i)]^2}$$

$$\nu_2 = \frac{-K\mathcal{E}(\delta_i u_i)}{\mathcal{E}(u_i)\mathcal{E}(\delta_i^2 u_i) - [\mathcal{E}(\delta_i u_i)]^2}$$

The maximum value of $(\sum p_i \delta_i q_i)^2 / \sum p_i u_i^{-1} q_i^2$ is $K^2/K\nu_1 = K/\nu_1$ which is

$$\mathcal{E}(\delta_i^2 u_i) - \frac{[\mathcal{E}(\delta_i u_i)]^2}{\mathcal{E}(u_i)} = \sum p_i \delta_i^2 u_i - \frac{(\sum p_i \delta_i u_i)^2}{(\sum p_i u_i)}$$

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 24	2. GOVT ACCESSION NO. AD-A120 059	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) When It Seems Desirable To Ignore Data		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Herman Chernoff		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0555
9. PERFORMING ORGANIZATION NAME AND ADDRESS Statistics Center, Massachusetts Institute of Technology, Cambridge, MA 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-609-001)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Statistics and Probability Code 436, Arlington, VA 22217		12. REPORT DATE Sept. 24, 1982
		13. NUMBER OF PAGES 42
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Hypothesis testing, optimal design, Pitman efficacy, Poisson

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See reverse side.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

An experiment designed to detect the relative motion of two astronomical objects raised the problem of testing, against shift alternatives, the hypothesis H_0 that two energy distributions are equivalent. The relevant data consist of independent Poisson counts X_{ij} with means $\lambda_j p_{ij} T_{ij}$ where λ_j is the intensity of radiation from the j-th object, p_{ij} is the probability that a random photon from the j-th object has energy in a small interval centered about e_i and T_{ij} is the time duration allocated to the count X_{ij}. The hypothesis H_0 implies that $P_{i1} = P_{i2}$ for $i = 1,2, \ldots, m$.

A natural test uses the statistic $\Sigma e_i(\hat{p}_{i2} - \hat{p}_{i1})$ where the $\hat{p}_{ij}$ are estimates of p_{ij}. For intervals where the p_{ij} were anticipated to be small, the experimenter chose small T_{ij} values and hence those $\hat{p}_{ij}$ were highly variable. Consequently, common sense suggests that the corresponding e_i, and X_{ij} be omitted in the above statistic, a practice which may be regarded as sinful by statistical dogma. This issue and others raised by the effects of small T_{ij} lead to the consideration of alternative test statistics and their relative efficiencies as well as the design problem of selecting T_{ij}.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

DATE
FILMED

11-82

DTIC